Astronomy & Astrophysics manuscript no. optimization˙xi
November 28, 2012

An optimized correlation function estimator for galaxy surveys

M. Vargas-Magaña[1], Julian. E. Bautista[1], J.-Ch. Hamilton[1], N.G. Busca[1], É. Aubourg[1], A. Labatie[2], J.-M. Le Goff[3], Stephanie Escoffier[4], Marc Manera[5], Cameron K. McBride[6], and Donald P. Schneider[7,8]
Christopher N. A. Willmer[9]

[1] APC, Astroparticule et Cosmologie, Université Paris Diderot, CNRS/IN2P3, CEA/Irfu, Observatoire de Paris, Sorbonne Paris Cité, 10, rue Alice Domon et Léonie Duquet, 75205 Paris Cedex 13, France
[2] Laboratoire AIM, CEA/DSM-CNRS-Université Paris Diderot, IRFU, SEDI-SAP, Service d'Astrophysique, Centre de Saclay, F-91191 Gif-Sur-Yvette cedex, France
[3] CEA centre de Saclay, irfu/SPP, F-91191 Gif-sur-Yvette, France
[4] CPPM, Aix-Marseille Université, CNRS/IN2P3, Marseille, France
[5] Institute of Cosmology and Gravitation, Portsmouth University, Dennis Sciama Building, Po1 3FX, Portsmouth, UK
[6] Harvard-Smithsonian Center for Astrophysics, 60 Garden St., Cambridge, MA 02138, USA
[7] Department of Astronomy and Astrophysics, The Pennsylvania State University, University Park, PA 16802,
[8] Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802
[9] Steward Observatory, University of Arizona 933 N. Cherry Avenue Tucson, AZ, 85721

Received xxxx / accepted xxxx

ABSTRACT

Measuring the two-point correlation function of the galaxies in the Universe gives access to the underlying dark matter distribution, which is related to cosmological parameters and to the physics of the primordial Universe. The estimation of the correlation function for current galaxy surveys makes use of the Landy-Szalay estimator, which is supposed to reach minimal variance. This is only true, however for a vanishing correlation function. We study the Landy-Szalay estimator when these conditions are not fulfilled and propose a new estimator that provides the smallest variance for a given survey geometry. Our estimator is a linear combination of ratios between pair-counts of data and/or random catalogues (DD, RR and DR). The optimal combination for a given geometry is determined by using log-normal mock catalogues. The resulting estimator is biased in a model dependent way, but we propose a simple iterative procedure to obtain an unbiased model independent estimator. Using various sets of simulated data (log-normal, second-order LPT and N-Body), we obtain a 20-25% gain on the error bars on the two-point correlation function for the SDSS geometry and ΛCDM correlation function. When applied on to SDSS data (DR7 and DR9), we achieve a similar gain on the correlation functions which translates in a 10-15% improvement on the estimation of the densities of matter, Ω_m, and dark energy, Ω_Λ in open LCDM model. The constraints derived from DR7 data with our estimator are similar to those obtained with the DR9 data and the Landy-Szalay estimator which covers a volume twice larger and with a density three times higher.

Key words. Cosmology – Large Scale Structure – Baryonic Acoustic Oscillations

Introduction

The distribution of galaxies in the Universe is an extremely rich source of information for cosmology since galaxies trace the underlying dark matter field. Studies of galaxy clustering at various redshifts therefore allow access to the expansion history of the Universe and, using models for this evolution, to constrain cosmology and fundamental physics [for example, see, Liddle and Lyth, (2000)]. On larger scales where fluctuations are still small, one can apply linear theory and have a direct access to cosmological parameters. On smaller scales, gravity acts in a nonlinear manner and the galaxy clustering allows one to investigate the structuration of dark matter into halos.

Observing the large-scale structure of the Universe is a promising approach for improving our understanding of its accelerated expansion observed by various cosmological probes in the last decade. The cosmic acceleration was initially proposed to reconcile the apparent low matter content of the Universe with a flat geometry in a standard Cold Dark Matter scenario [Efstathiou et al., (1990)]. The first convicing measurement of cosmic acceleration came from observations that type Ia supernovae appeared less luminous than expected in a decelerating Universe [Riess et al., (1998), Perlmutter et al., (1999)]. These observations can be accomodated by modifying General Relativity on cosmological scales or, within a Friedmann-Lemaître-Robertson-Walker (FLRW) cosmology, by adding a *Dark Energy* component with a density $\Omega_X \sim 0.7$, a negative pressure, and a possibly evolving equation of state. Since then, the cosmic acceleration has been confirmed by other probes with abundant data, including Cosmic Microwave Background (CMB) fluctuations [Komatsu et al., (2011), Sherwin et al., (2011)], Integrated Sachs–Wolfe (ISW) effect [Granett et al., (2009)] and Baryonic Acoustic Oscillations (BAO) (Weinberg et al., (2012) for a general review and Anderson et al., (2012) for the latest measurement). These data point towards a Dark Energy with a constant equation-of-state parameter, $w = -1$,

or equivalently a pure cosmological constant. Baryonic Acoustic Oscillations measurements are based on the observation of an acoustic peak in the correlation function of the matter density fluctuations, corresponding to the acoustic horizon at the epoch of matter-radiation decoupling [Eisenstein and Hu, (1998)]. The acoustic scale is used as a standard ruler at various redshifts, allowing for the measurement of the angular distance in the transverse directions and the expansion rate in the radial direction [Reid et al.; 2012].

When investigating the large-scale structure of the Universe using galaxies as a tracer of dark matter, one needs large-field-of-view deep galaxy surveys such as Sloan Digital Sky Survey (SDSS-III) Baryon Oscillation Spectroscopic Survey (BOSS) [Eisenstein et al., (2011)], i.e. high density galaxy catalogues, where the radial positions of galaxies are measured by their redshifts. The two-point correlation function is commonly used for characterizing the large scale structure within such galaxy surveys. The fact that one does not directly measure the density within the survey volume, but samples this density through galaxy locations, makes the estimation of the two-point correlation function more complex. The observed quantity is the average number of neighbors at a given distance in the survey volume and is biased by the fact that galaxies near the edges of the catalogue volume have less neighbors than they should have, which needs to be corrected for in an optimal way. This issue does not occur, for example when directly measuring a function of the matter density through the Lyman-α forest of distant quasars [Slosar et al., (2011)].

In this article, we introduce a novel estimator for the two-point correlation function of galaxies. Its performance can be optimized for a given galaxy survey geometry. In section 1 we motivate this effort, showing that various well-known estimators for the two-point correlation function have a bias and a variance that strongly depend on the survey geometry. The commonly used Landy-Szalay estimator [Landy and Szalay, (1993)] has been shown to be both unbiased and of minimal variance in the limit of a vanishing correlation function. We show that in realistic cases, where the correlation function is not zero, the Landy-Szalay estimator does not reach the Poisson noise limit. For pedagogical reasons, we start in section 2 with a simpler but biased version of our optimal estimator and we develop in section 3 a simple iterative procedure that allows the final estimator to be model independent, with an improvement of the accuracy around 20-25% with respect to the Landy-Szalay estimator. In section 4 we apply our final estimator to data from the SDSS-II Seventh Data Release (DR7) Luminous Red Galaxy sample and on the SDSS-III/BOSS DR9 "CMASS " sample and show the improvement on the two-point correlation function measurement and cosmological parameters with respect to previous analyses.

1. Motivations for an optimized two-point correlation function estimator

1.1. Commonly used estimators

Estimators of the two-point correlation function $\xi(s)$ (s being the comoving separation) have been studied by various authors [Peebles and Hauser, (1974), Davis and Peebles, (1983), Hewett, (1982),Hamilton, (1993),Landy and Szalay, (1993)]. Generically, pair counts in data are compared to pair counts in random samples that follow the geometry of the survey. Let us assume a catalogue of n_d objects in the data sample and n_r in the random sample and calculate three sets of numbers of pairs as a function of the binned comoving separation s[1]:

- within the data sample, leading to $dd(s)$ that can be normalized to the total number of pairs as: $DD(s) = \dfrac{dd(s)}{n_d(n_d-1)/2}$.
- within the random sample, leading to $rr(s)$ normalized as: $RR(s) = \dfrac{rr(s)}{n_r(n_r-1)/2}$.
- among both samples (cross correlation) leading to $dr(s)$ normalized as: $DR(s) = \dfrac{dr(s)}{n_r n_d}$.

The most common estimators discussed in the literature are:

- $\hat{\xi}_{PH}(s) = \dfrac{DD}{RR} - 1$ [Peebles and Hauser, (1974)]
- $\hat{\xi}_{Hew}(s) = \dfrac{DD - DR}{RR}$ [Hewett, (1982)]
- $\hat{\xi}_{DP}(s) = \dfrac{DD}{DR} - 1$ [Davis and Peebles, (1983)]
- $\hat{\xi}_{H}(s) = \dfrac{DD \times RR}{DR^2} - 1$ [Hamilton, (1993)]
- $\hat{\xi}_{LS}(s) = \dfrac{DD - 2DR + RR}{RR}$ [Landy and Szalay, (1993)]

Some studies have compared the behavior of the different two-correlation-function estimators, mainly in the small-scale regime and using smaller samples. In Pons-Bordería et al. (1999), 6 estimators were analyzed, including both the Hamilton and Landy-Szalay estimators, and the authors did not find any outstanding winner among those estimators. In Kerscher (1999) and Kerscher, Szapudi and Szalay (2000), 9 estimators were considered and the estimators presenting the best properties were the Landy-Szalay and Hamilton estimators.

1.2. Relative performances of the common estimators

To compare the performances of these estimators, we have used two sets of 120 mock catalogues obtained from lognormal [Coles and Jones, (1991)] density field simulations containing about 271,000 galaxies in both a cube of 1 h^{-1} Gpc size and a far more complex geometry corresponding to the BOSS (DR9) survey [Anderson et al., (2012)] which contains roughly the same volume as the cube. In addition we used random catalogues with three as many galaxies than the mock catalogues for both geometries. The cosmology used for the lognormal fields is taken from the WMAP 7 years analysis [Komatsu et al., (2011)].

Fig. 1 shows the correlation function obtained with the different estimators for the cubic and DR9 geometries. We

[1] The number of pairs can be spherically averaged in the simplest approach. Its dependance on the angle with respect to the line of sight can be considered in a more elaborated analysis, in order to account for the sensitivity to angular distance in the transverse direction and $H(z)$ in the radial one (see [Cabré and Gaztañaga, (2008)] for details).

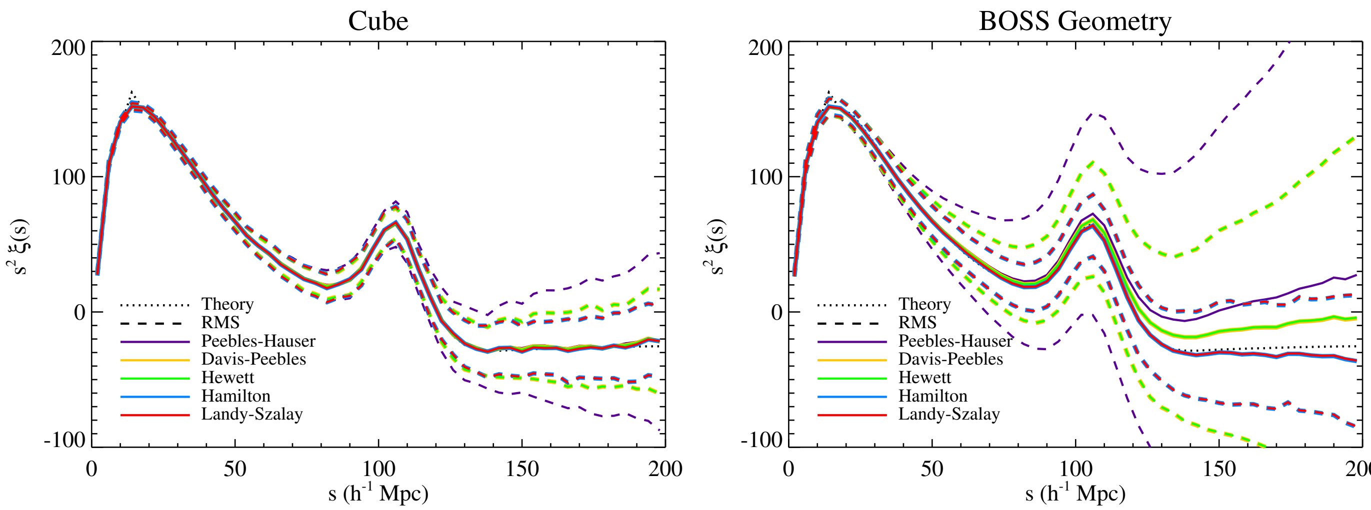

Fig. 1. Input two-point correlation function (dotted black line) and reconstructed ones using the various estimators available in the literature (solid lines of various colors) for a cubic geometry (left) or a realistic (BOSS DR9) survey volume (right). The dashed lines of various colors represent the RMS of the corresponding estimators. The Hamilton and Landy-Szalay lines are exactly superposed as well as the Davis-Peebles and Hewett lines. (Coloured version of the figure is available online)

clearly see differences between the performances of the estimators in the cube and in the DR9, either their mean result (solid lines) or their root-mean-square errors, RMS (dashed lines). In the case of the DR9 geometry, the mean result obtained with the Peebles-Hauser, Davis-Peebles and Hewett estimators is more biased with respect to the theory at large scales. Landy-Szalay and Hamilton estimators are much less biased than the others in this more complex geometry. Examining at the RMS, all estimators have their accuracy degraded by the effects of geometry. Landy-Szalay and Hamilton again show best performances with the lowest variances in both geometries as expected.

1.3. Optimality of the Landy-Szalay estimator

In the limit of an infinitely large random catalogue, for which the volume is much larger than the observed scales, and a vanishing two-point correlation function (uniform galaxy distribution), the Landy-Szalay estimator is known to be unbiased and of minimal variance. It is therefore the most widely used in modern galaxy surveys [e.g., Eisenstein et al., (2005), Percival et al, (2007),Kazin et al., (2010),Blake et al., (2011), Anderson et al., (2012), Sanchez et al., (2012)]. In practice the volume of modern survey is sufficiently large and one can also produce a large enough random catalogue, but the correlation function to be measured is nonzero. So it is crucial to check bias and variance of estimators in case of realistic non zero correlation functions.

Using additional lognormal simulations, we have investigated the RMS of the Landy-Szalay estimator as a function of the size of the random catalogue for both a zero correlation function and the one expected from the ΛCDM scenario. Fifty realizations were produced in both cases where a cubic geometry was used in order to be insensitive to the degradation due to the survey geometry. The resulting RMS are shown in Fig. 2, along with the expectations for an optimal estimator (from equation 48 in Landy and Szalay, (1993) accounting for the finite size of the random catalogue. It appears that, when the correlation function is not vanishing, the Landy-Szalay estimator does not reach the Poisson noise limit. This suggests that a better estimator can be found in the case of a nonvanishing correlation function and a more complicated survey geometry.

2. An optimized estimator

2.1. General form and optimization criterion

Our search for a better estimator started from the observation that the commonly used estimators are linear combinations of ratios of pair counts, $\boldsymbol{DD}$, $\boldsymbol{DR}$ and $\boldsymbol{RR}$ (hereafter the s dependence is described by vectors), with the exception of the Hamilton estimator, which involves ratios of second order products of pair counts. We therefore investigate an estimator which would be an optimal linear combination of all possible ratios $\boldsymbol{R}_i$ up to second order. Table 1 summarizes the six ratios at first order and the twelve at second order. The generic optimal estimator can then be expressed as:

$$\hat{\boldsymbol{\xi}}^{opt}(\boldsymbol{c}) = c_0 + \sum_{i=1}^{6} c_i \boldsymbol{R}_i^{(1)} + \sum_{i=7}^{18} c_i \boldsymbol{R}_i^{(2)} \,. \tag{1}$$

The nineteen c_i coefficients are optimized in order to minimize the variance of the estimator for a given geometry. This optimization is done through a χ^2 minimization using a large set of mock catalogues generated using lognormal fields, for which DD, DR and RR are stored, so that all the R_i terms can be calculated. The χ^2 is minimized with respect to the vector of parameters $\boldsymbol{c}$ as:

$$\chi^2 = \sum_j \left[\hat{\boldsymbol{\xi}}_j^{\rm opt}(\boldsymbol{c}) - \boldsymbol{\xi}_{\rm th}\right]^t \cdot N_{\rm LS}^{-1} \cdot \left[\hat{\boldsymbol{\xi}}_j^{\rm opt}(\boldsymbol{c}) - \boldsymbol{\xi}_{\rm th}\right] \;, \tag{2}$$

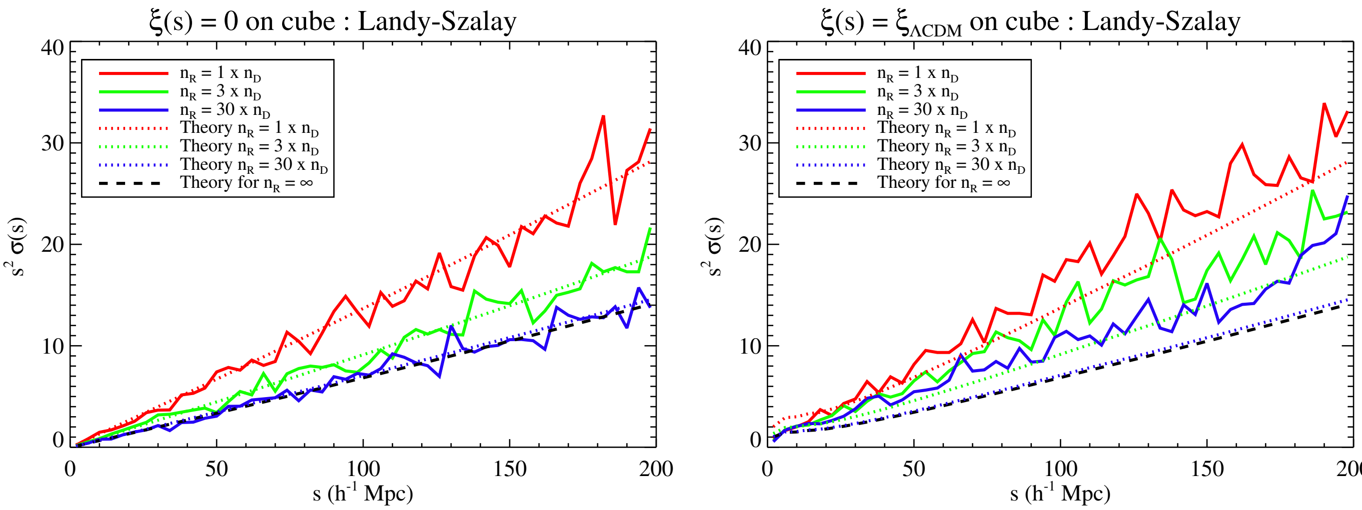

Fig. 2. RMS of the Landy-Szalay estimator for lognormal simulations in a cubic geometry and for either a zero two-point correlation function (left) or the ΛCDM expectation (right). The RMS is shown in different colors for various sizes of the Random sample relative to the data sample. A dotted line of the same color shows the expectation for an optimal estimator in each case. Finally, the ultimate limit, corresponding to an infinite size of the random sample, is shown as a black dashed line. (Coloured version of the figure is available online)

Table 1. The nineteen ratios formed by using pair counts up to second order.

0th order		
1		
1st order terms $R^{(1)}$		
$\frac{DD}{RR}$	$\frac{DR}{RR}$	$\frac{DR}{DD}$
$\frac{RR}{DD}$	$\frac{RR}{DR}$	$\frac{DD}{DR}$
2nd order terms $R^{(2)}$		
$\frac{DR\times RR}{DD^2}$	$\frac{RR^2}{DD^2}$	$\frac{DR\times DD}{RR^2}$
$\frac{DD^2}{RR^2}$	$\frac{DR^2}{RR^2}$	$\frac{DD^2}{DR^2}$
$\frac{RR^2}{DR^2}$	$\frac{DD\times RR}{DR^2}$	$\frac{RR^2}{DD\times DR}$
$\frac{DR^2}{DD\times RR}$	$\frac{DD^2}{DR\times RR}$	$\frac{DR^2}{DD^2}$

where the j index stands for the j-th realization, $\hat{\boldsymbol{\xi}}^{\mathrm{opt}}$ is the vector of the values of the estimator in the comoving distance s bins and $\boldsymbol{\xi}_{\mathrm{th}}$ the vector for the theoretical input correlation function. The quantity N_{LS} is the covariance matrix of fluctuations of $\boldsymbol{\xi}_j^{\mathrm{LS}}$ around the mean Landy-Szalay correlation function $\langle\boldsymbol{\xi}^{\mathrm{LS}}\rangle$, the mean taken over the mock realizations.

This approach will result in an estimator with a variance at most as large as that of the Landy-Szalay estimator, but which might have a significant bias. This bias can be calculated and corrected for to an arbitrary precision for a given input correlation function (therefore for a given cosmological model). However, this bias correction is model dependent and would only work perfectly when the input cosmology in the mock data matches the one to be measured in the real data. In section 3 we propose a simple iterative method that allows efficient circumvention of this problem.

2.2. Performances on simulations

Using lognormal simulations, we produced 120 realizations of galaxy catalogues with a geometry similar to that of the SDSS-III/BOSS (DR9) survey [Eisenstein et al., (2011), Anderson et al., (2012)]. The fiducial cosmology was defined by $h = 0.7$, $\Omega_m = 0.27$, $\Omega_\Lambda = 0.73$, $\Omega_b = 0.045$, $\sigma_8 = 0.8$ and $n_s = 1.0$. For each realization, we generate a random catalogue with the same geometry and calculate $\boldsymbol{DD}$, $\boldsymbol{DR}$ and $\boldsymbol{RR}$ for comoving separations between 0 and 200 h^{-1} Mpc with bins of 4 h^{-1} Mpc. We then calculate the Landy-Szalay estimator for each simulation $\boldsymbol{\xi}_j^{\mathrm{LS}}$, the average estimator $\langle\boldsymbol{\xi}^{\mathrm{LS}}\rangle$, and its covariance matrix N_{LS}, empirically, i.e., from the dispersion of the individual realizations.

We then have all the ingredients required to minimize the χ^2 in Eq. 2 and obtain an optimal estimator, which was done by limiting the χ^2 to the region [40, 200] h^{-1} Mpc. The actual value of the coefficients c_i is actually not particularly meaningful for two reasons: it depends on the geometry of the survey and is therefore not "general"; in addition, the parameters are degenerate because the nineteen $\boldsymbol{R}_i$ terms are not independent.

Fig. 3 compares our estimator to the Landy-Szalay estimator. The residual with respect to the theoretical input correlation and the RMS are shown for both estimators. These RMS are just the square-roots of the diagonal elements of the estimator covariance matrix, calculated empirically from the individual realizations. The Landy-Szalay

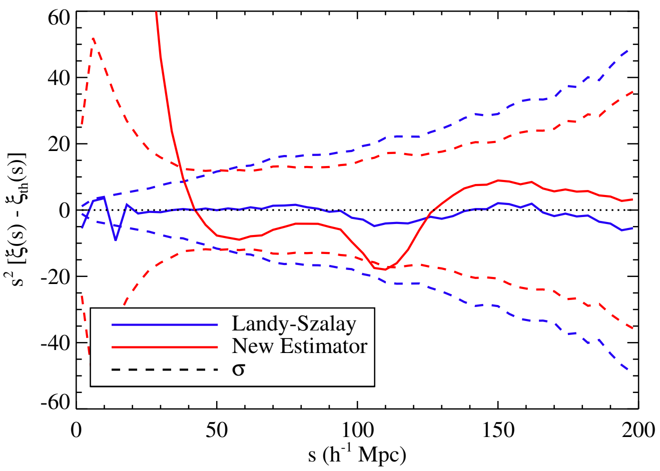

Fig. 3. Residuals (solid) and RMS (dashed) of the Landy-Szalay estimator (blue) and of the estimator fitted to minimize the variance (red) as explained in the text. (Coloured version of the figure is available online)

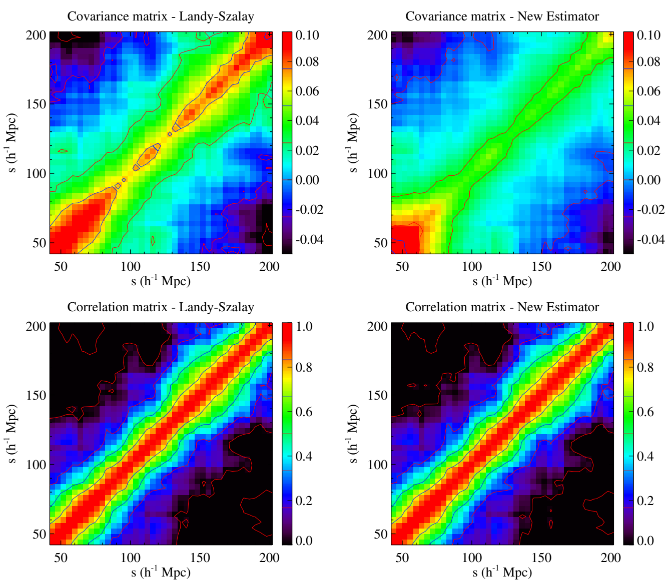

Fig. 4. [Top panels] Covariance matrices multiplied by the square of the comoving distance. [Bottom panels] Correlation matrices for both the Landy-Szalay estimator (left panels) and the optimized estimator (right panels). (Coloured version of the figure is available online)

estimator is essentially unbiased while a significant bias is observed for our estimator, which, however, remains within the 1σ range. In contrast, our optimized estimator appears to have smaller variances than the Landy-Szalay estimator in the region [40, 200] h^{-1} Mpc , where the fit was performed. Fig. 4 suggests that the covariance and correlation matrices for the Landy-Szalay and optimized estimators. The latter has a smaller covariance matrix and no extra correlation between the bins.

2.3. Model dependance

Fig. 3 and Fig. 4 show that by correcting the optimized estimator by its average bias, which can be known with excellent accuracy by having a large number of mock realizations, one can achieve a better accuracy on the correlation function than the Landy-Szalay estimator. Unfortunately, the bias exhibits a peak at the location of the BAO scale and therefore will be different in another cosmology: it is strongly model dependent. If one uses an estimator optimized with a set of simulations that assumed a cosmology different from the actual one, the peak position in the bias will be different from that in the data, resulting in a strong distortion of the peak shape after bias correction and in a shift in its location. This is illustrated in Fig. 5, which will be further discussed latter. Fortunately, one can eliminate the cosmology dependence of the fitting, as described in next section.

3. Iterative optimized estimator

To transform the optimized estimator into a model independent one, we investigated the possibility of iterating with an estimator that assumes the same cosmology as that derived from the data. Such a procedure could be quite time consuming, as one needs a large number of mock realizations for a given cosmology to optimize the estimator for this cosmology. We have found a way to do this efficiently, limiting the number of simulations to a few times the initial one.

3.1. Description of the method

Our iterative procedure starts with a first calculation of the correlation function using the Landy-Szalay estimator. We then fit the resulting correlation function with a model that has considerable freedom on the general broadband shape, so that it is essentially sensitive to the location of the acoustic peak, as used in BOSS analysis [Anderson et al., (2012)]:

$$\xi_{\rm data}(s) = b^2\ \xi_{\rm theory}(\alpha s) + a_0 + \frac{a_1}{s} + \frac{a_2}{s^2}\ , \tag{3}$$

where $\xi_{\rm theory}$ is the theoretical linear model from Eisenstein and Hu, (1998), b is the constant galaxy dark-matter bias factor and a_0, a_1 and a_2 are nuisance parameters.

From this fit, we obtain the first iteration of the dilation scale parameter, α, that characterizes the location of the peak:

$$\alpha = \left(\frac{D_V}{r_s}\right) / \left(\frac{D_V}{r_s}\right)_f\ , \tag{4}$$

where r_s is the comoving sound horizon at decoupling; the subscript f means that the quantity is calculated using our fiducial cosmology, for which $r_s = 157.42$ Mpc; $D_V(z)$ is the spherically averaged distance to redshift z and is defined by [Mehta et al., (2011)]:

$$D_V(z) = \left((1+z)^2 \frac{D_A^2(z) cz}{H(z)}\right)^{1/3}\ . \tag{5}$$

The parameter α is unity if the actual cosmology matches the fiducial one. The result of this fit is a first

estimate of the cosmological model suggested by the data, labeled by α_0. This is actually the result of the standard analysis with the Landy-Szalay estimator.

We then perform a large number of realizations of mock catalogues with the same DR9 geometry as the data, for various values of α around α_0. We use lognormal simulations that can be quickly generated. For this work we have simulated 120 realizations of 9 different cosmologies such that the dilation parameter α covers the range $[0.96, 1.04]$ in steps of width 0.01. For each realization we used the same number of galaxies, about 271,000, and a random catalogue 15 times larger. This step does not require more than a few days on desktop machines.

For the set of 120 simulations corresponding to a given input α_k, one can find the coefficients $\boldsymbol{c}_k$ by minimizing the χ^2 defined in Eq. 2. The resulting correlation function, $\hat{\boldsymbol{\xi}}_j(\boldsymbol{c}_k)$, for the realization j is given by Eq. 1. We compute the average bias $\boldsymbol{B}_k$ of the correlation function with respect to the theory $\boldsymbol{\xi}_{\rm theory}(\alpha_k)$:

$$\boldsymbol{B}_k = \left\langle \hat{\boldsymbol{\xi}}_j(\boldsymbol{c}_k) - \boldsymbol{\xi}_{\rm theory}(\alpha_k) \right\rangle_j . \tag{6}$$

The covariance matrix $N_{\rm opt}(\alpha_k)$ is obtained from the fluctuations of the same 120 realizations :

$$N_{\rm opt}(\alpha_k) = \left\langle \left[\boldsymbol{\xi}_j(\boldsymbol{c}_k) - \bar{\boldsymbol{\xi}}(\boldsymbol{c}_k)\right] \left[\boldsymbol{\xi}_j(\boldsymbol{c}_k) - \bar{\boldsymbol{\xi}}(\boldsymbol{c}_k)\right]^T \right\rangle_j . \tag{7}$$

Hereafter we redefine the process of applying the estimator corresponding to α_k to a data sample by two steps:

- use Eq. 1 with coefficients $\boldsymbol{c}_k$ to calculate $\hat{\boldsymbol{\xi}}_{\rm data}(\boldsymbol{c}_k)$
- add the bias of the estimator, $\boldsymbol{B}_k$.

We can now proceed with the iterative procedure. Since the first iteration value, $\alpha = \alpha_0$, is not exactly one of the nine available α_k, we apply the estimator corresponding to the closest two values of α_0, $\alpha_{\rm lo}$ and $\alpha_{\rm hi}$, to the data, and we interpolate between the two resulting correlation functions:

$$\boldsymbol{\xi}^{\rm opt}(\alpha_0) = (1-t)\ \boldsymbol{\xi}(\alpha_{\rm lo}) + t\ \boldsymbol{\xi}(\alpha_{\rm hi}) , \tag{8}$$

where $t = (\alpha_0 - \alpha_{\rm lo})/(\alpha_{\rm hi} - \alpha_{\rm lo})$. Similarly, the covariance matrix can be written as a function of the two covariance matrices $N_{\rm opt}(\alpha_{\rm lo})$ and $N_{\rm opt}(\alpha_{\rm hi})$ as:

$$\begin{aligned} N_{\rm opt}(\alpha) = &(1-t)^2 N_{\rm opt}(\alpha_{\rm lo}) + t^2 N_{\rm opt}(\alpha_{\rm hi}) \\ &+ t(1-t)\ C_{\rm opt}(\alpha_{\rm lo}, \alpha_{\rm hi}) , \end{aligned} \tag{9}$$

where $C_{\rm opt}(\alpha_{\rm lo}, \alpha_{\rm hi})$ is the cross-covariance between $\boldsymbol{\xi}^{\rm opt}(\alpha_{\rm lo})$ and $\boldsymbol{\xi}^{\rm opt}(\alpha_{\rm hi})$ given by:

$$\begin{aligned} &C_{\rm opt}(\alpha_{\rm lo}, \alpha_{\rm hi}) = \\ &\Big\langle \left[\boldsymbol{\xi}^{\rm opt}(\alpha_{\rm lo}) - \bar{\boldsymbol{\xi}}^{\rm opt}(\alpha_{\rm lo})\right] \left[\boldsymbol{\xi}^{\rm opt}(\alpha_{\rm hi}) - \bar{\boldsymbol{\xi}}^{\rm opt}(\alpha_{\rm hi})\right]^T \\ &+ \left[\boldsymbol{\xi}^{\rm opt}(\alpha_{\rm hi}) - \bar{\boldsymbol{\xi}}^{\rm opt}(\alpha_{\rm hi})\right] \left[\boldsymbol{\xi}^{\rm opt}(\alpha_{\rm lo}) - \bar{\boldsymbol{\xi}}^{\rm opt}(\alpha_{\rm lo})\right]^T \Big\rangle_s . \end{aligned} \tag{10}$$

Finally, we fit the correlation function (Eq. 8) with the template (Eq. 3) using the covariance matrix (Eq. 9), which yields a new value α_1 at the second iteration. We then iterate until the estimated α_i varies less than a given quantity ($\Delta\alpha = 0.0001$) between two successive iterations. In practice, convergence is achieved after a few iterations.

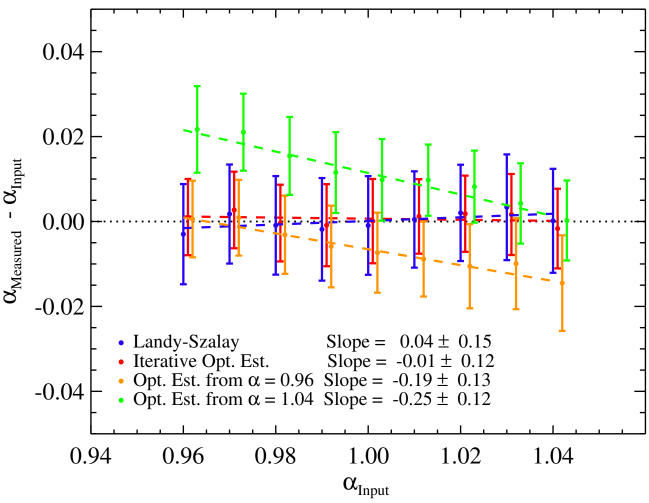

Fig. 5. Difference between the mean $\alpha_{\rm Measured}$ and $\alpha_{\rm Input}$ as a function of $\alpha_{\rm Input}$ for lognormal simulations, when applying four estimators: the Landy-Szalay estimator (blue points), the Iterative Optimal Estimator (red points), and two non-iterative estimators with $\alpha = 0.96$ (yellow points) and 1.04 (green points). Dashed lines of the same colors, fitted to the points, emphasize the biases. (Coloured version of the figure is available online)

3.2. Performance on simulations

In this section, we investigate the properties of the iterative optimal estimator on mock catalogs. We start with the lognormal simulations used to optimize the estimator and show that we derive an estimator that is indeed independent of the input cosmological model. We obtain an 20-25% increase in the accuracy on the α parameter with these simulations. We also test our estimator on other simulations than the lognormal ones used to optimize it. These are more realistic simulations than the lognomal simulations; they were produced in the framework of the SDSS-III/BOSS galaxy clustering working groups.

3.2.1. Lognormal mock data

As an illustration, we first considered what happens when we do not use the iterative procedure. The nine different sets of 120 lognormal simulations provide nine different optimal estimators, defined by $\boldsymbol{c}_k$ and $\boldsymbol{B}_k$. We choose two of them to apply to the nine sets of simulations *without* the iterative procedure. We fit the resulting correlation function with the template of Eq. 3 to obtain the scale parameter, $\alpha_{\rm Measured}$. In Fig. 5, $\alpha_{\rm Measured}$ is averaged over the 120 simulations with the same given $\alpha_{\rm Input}$, and its difference with $\alpha_{\rm Input}$ is plotted versus $\alpha_{\rm Input}$. The non-iterative estimators do not recover $\alpha_{\rm Input}$. This result occurs because the peak-shaped (Fig. 3) bias correction, $B(\alpha, r)$ slightly shifts the peak position of the data to the left ($\alpha = 0.96$) or to the right ($\alpha = 1.04$). As emphasized by the linear fits in the figure, the bias increases with the difference $\alpha - \alpha_{\rm Input}$.

Fig. 5 also demonstrates what happens with the iterative optimal estimator. The iterative procedure indeed removes the bias, since the iterative optimal estimator appears to be unbiased. The error bars in the figure are RMS

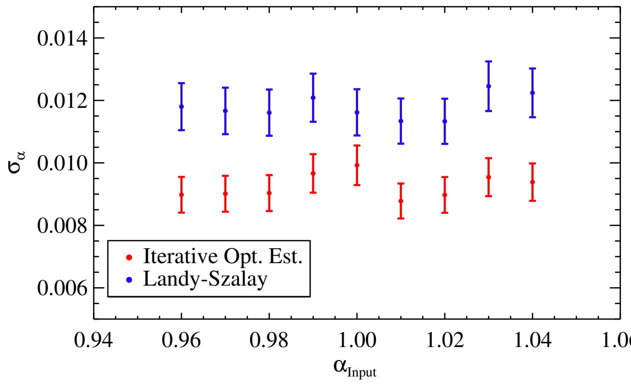

Fig. 6. The error on $\alpha_{\rm Measured}$ for the iterative optimal (red points) and Landy-Szalay (blue points) estimators for the nine sets of realizations with different $\alpha_{\rm Input}$. (Coloured version of the figure is available online)

of the values of $\alpha_{\rm Measured}$ for the 120 different simulations. The optimal estimator gives smaller RMS than the Landy-Szalay estimator. This result is confirmed in Fig. 6, which shows this RMS as a function of $\alpha_{\rm Input}$. The gain obtained with the optimal estimator is ≈22% relative to the Landy-Szalay estimator (this number is not a general one; the precise value depends on the geometry of the survey) leading to similar improvement on subsequent cosmological constraints.

3.2.2. PTHalos and LasDamas Mock data

The studies in the previous section were performed by applying the iterative optimal estimator to the lognormal simulations that were used to optimize the estimator. We repeated the calculations using two other sets of mock catalogues that have very similar geometries, based on the BOSS DR9 footprint [Anderson et al., (2012)].

The first set is based on 2nd order Lagrangian Perturbation Theory matter field [Scoccimaro and Sheth, (2002)] and halo occupation function named PTHalos [Manera et al., (2012)]. A total of 610 realizations were produced with $h = 0.7$, $\Omega_m = 0.274$, $\Omega_\Lambda = 0.726$, $\Omega_b h^2 = 0.0224$, $\sigma_8 = 0.8$ and $n_s = 0.97$. [2]. As the fiducial cosmology used to compute the comoving distances of the galaxies is slightly different than the one used in mock catalogues, α (Eq. 4) is not expected to be 1 but 1.002.

The second set is even more realistic; it uses N-body simulations, named Large Suite of Dark Matter Simulations (LasDamas) [McBride et al., in preparation]; developed within the SDSS I-II galaxy clustering working group for the DR7 LRG analysis. A total of 153 realizations were produced assuming a flat ΛCDM cosmology with $\Omega_b = 0.04$, $\Omega_m = 0.25$, $h = 0.7$, $n_s = 1.0$ and $\sigma_8 = 0.8$, for which α is expected to be 0.988.

Fig. 7 shows the "pull" histogram of the correlation functions, i.e., the residuals of the correlation functions relative to the average Landy-Szalay correlation function, normalized to the empirical RMS of the Landy-Szalay estimator, $(\xi - \langle \xi_{LS} \rangle)/\sigma_{LS}$. By construction, the width of the pull

[2] For this work we used 598 of the 610 realizations available.

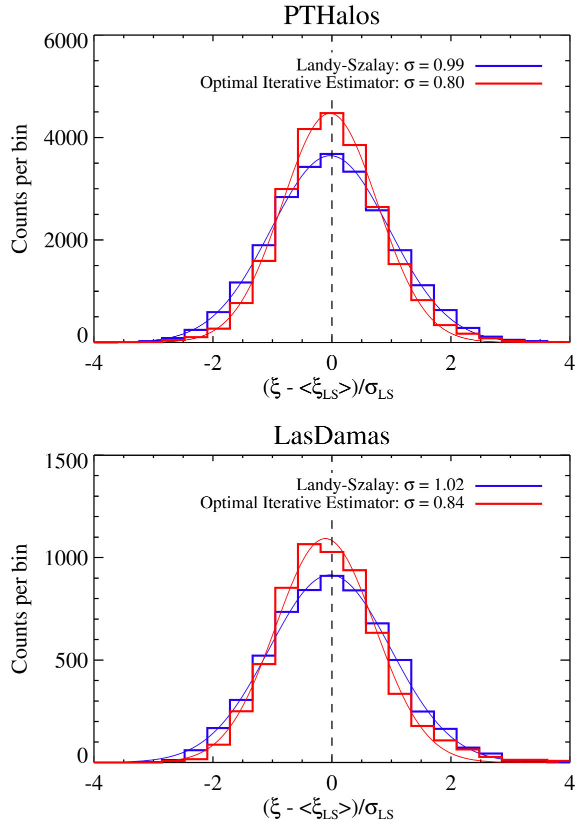

Fig. 7. "Pull" distribution of correlation functions measured with PTHalos (top) and LasDamas (bottom) mocks in the range $40 < s < 200\ h^{-1}$ Mpc with the Landy-Szalay (blue) and the Iterative Optimal estimator (red). The standard deviation of the Gaussian fit shows a smaller scatter for the latter estimator. (Coloured version of the figure is available online)

distribution for the Landy-Szalay estimator is close to 1 for both sets of mock data, while for the iterative optimal estimator it is 0.80 (PTHalos) and 0.84 (LasDamas), which is similar to the 22% gain on the error bars obtained with the lognormal simulations.

This result is confirmed by Fig. 8, which shows the covariance and correlation matrices obtained with both estimators on the PTHalos mocks. The gain in the covariance matrix elements is obvious and is not mitigated by an increase of the off-diagonal terms in the correlation matrix. Fig. 9 shows the same information for the LasDamas simulations. The matrices are noisier since we have a smaller number of realizations, but the improvement is visible, and again the correlation does not change. A small increase of the covariance matrix is present between 40 and 60 h^{-1} Mpc, which can be also seen in Fig. 3. However, these scales are much smaller than the scales of interest here (BAO peak) where we indeed see a clear reduction of the covariance.

Finally, Fig. 10 displays the improvement on the estimation of α obtained using the iterative optimal estimator. The scatter of $\alpha_{\rm Measured}$ with the optimal estimator is reduced relative to Landy-Szalay by 21% for PTHalos and 17% for LasDamas mocks. These gains are consistent with the observed "pull" distribution (Fig. 7) and confirme the gain observed with the lognormal simulations.

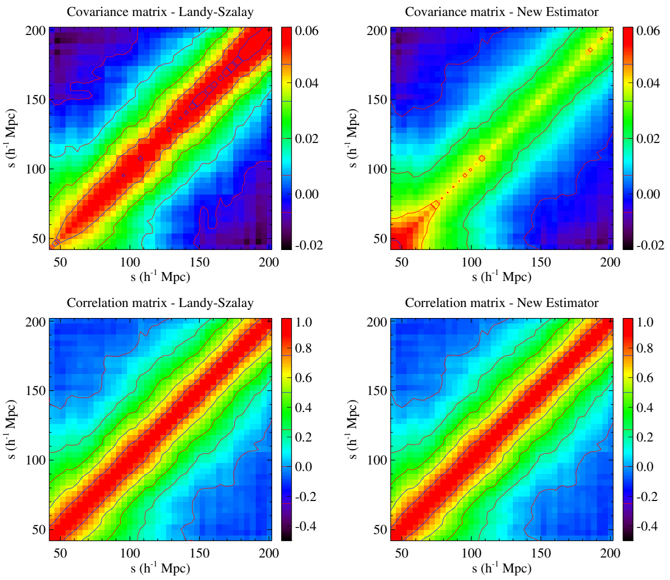

Fig. 8. Covariance matrices times the square of the comoving distance (top panels) and correlation matrices (bottom panels) of PTHalos mock catalogues using Landy-Szalay (left panels) and the Iterative Optimal Estimator (right panels). (Coloured version of the figure is available online)

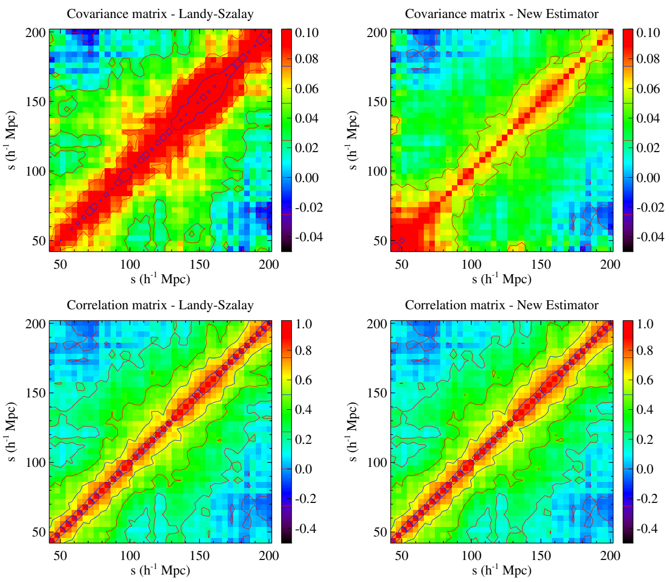

Fig. 9. Same as figure 8 for LasDamas mock catalogues. (Coloured version of the figure is available online)

4. Application to real data

4.1. Data description

We apply our final estimator on two galaxy samples: the SDSS I-II DR7 Luminous Red Galaxy sample (LRG) [Eisenstein et al., (2001)] and the SDSS-III/BOSS DR9 CMASS [Padmanabhan et al. (2012)].

Both surveys, SDSS-I-II and SDSS-III/BOSS use the same wide field, dedicated telescope, the 2.5 m-aperture Sloan Foundation Telescope at Apache Point Observatory in New Mexico [Gunn, et al. (2006)]. Those surveys imaged the sky at high latitude in the ugriz bands [Fukugita et al. (1996)], using a mosaic CCD camera with a field of view spanning 2.5 deg [Gunn, et al. (1998)]. The SDSS-I-II imaging survey is described in Abazajian et al., (2009). This galaxy catalogue had been built based on the prescription in Eisenstein et al., (2001), selecting the most luminous galaxies since they are more massive and then more biased with respect to the dark-matter density field. More details about the construction of the catalogue can be found in Kazin et al., (2010). The galaxies in the LRG sample have redshifts in the range $0.16 < z < 0.47$ and a density of about $10^{-4}h^3$ Mpc^{-3}.

The BOSS imaging survey data are described in Aihara et al. (2011), the spectrograph design and performance in Smee, S.A., et al., (2012), and the spectral data reductions in Schlegel et al. (2012) and Bolton, A., et al. (2012). A summary of BOSS can be found in Dawson, K., et al. (2012). The SDSS Data Release 9 [Ahn et al. (2012)] CMASS sample of galaxies is constructed using an extension of the selection algorithm of DR7 LRG sample in order to detect fainter and bluer massive galaxies lying in the redshift range $0.43 < z < 0.7$. The final density of this sample is $3 \times 10^{-4}h^3$ Mpc^{-3}. A more detailed explanation of the target selection is given in Padmanabhan et al. (2012).

4.2. DR7

We compared the iterative optimal estimator to the Landy-Szalay estimator for the estimation of the spherically averaged 2-point correlation function of the SDSS DR7 [Abazajian et al., (2009)] LRG sample (Fig. 11). In order to estimate the correlation function, we used a random catalogue 15 times larger than the data sample. The coefficients $\boldsymbol{c}_k$ and biases $B_k(\boldsymbol{r})$ for the iterative estimator were obtained using the nine sets of lognormal simulations as described in section 3.1. The covariance matrix of the data correlation function for both estimators comes from the 153 realizations of the LasDamas mocks. The left panel of Fig. 11 shows the resulting correlation function for both estimators. The error bars obtained for the iterative estimator are smaller than for the Landy-Szalay estimator, but both curves are consistent with each other.

The correlation functions were fitted as for the mock catalogues, using the template defined by Eq. 3. The resulting values of $\alpha_{\rm Measured}$ are compatible with unity and the error for the optimal estimator is lower by 31%, as shown in Table 2. This improvement is larger than the 17% improvement on the mean error observed on mock data but it is consistent with the scatter of the errors (Fig. 13, left).

In Fig. 12 (left) we use both estimators to compare $\alpha_{\rm Measured}$ for the DR7 LRG sample (red points) and LasDamas realizations (black points). The DR7 measurement is well inside the LasDamas cloud, being very close to the mean.

Another way to improve the measurement accuracy of the BAO peak is through the reconstruction technique, where galaxies are slightly displaced so that the density field is as it should be without non-linear structure growth effects [Eisenstein et al., (2007)]. Xu et al., (2012) used the reconstruction technique on the DR7 LRG sample. Before reconstruction they obtain $\alpha = 1.015 \pm 0.044$; after reconstruction $\alpha = 1.012 \pm 0.024$, an improvement of 45%. Our estimator, with an 31% improvement, yields $\alpha = 1.006 \pm 0.018$, consistent with the reconstruction result. This comparison shows

Table 2. Values of α found with two different estimators of the correlation function for each sample.

Sample	Landy-Szalay $\alpha_{\rm LS}$	It. Opt. Est. $\alpha_{\rm opt}$	Gain %
Mean LasDamas	0.976 ± 0.035	0.979 ± 0.029	17
Mean PTHalos	1.013 ± 0.039	1.011 ± 0.031	21
DR7	1.004 ± 0.026	1.006 ± 0.018	31
DR9	1.010 ± 0.018	1.009 ± 0.013	28

Table 3. Improvement on cosmological parameters with the iterative optimal estimator.

WMAP7+	Ω_m	Gain	Ω_Λ	Gain
DR7 (LS)	0.276 ± 0.018	-	0.727 ± 0.017	-
DR7 (It. Opt.)	0.274 ± 0.014	22%	0.729 ± 0.014	17%
DR9 (LS)	0.278 ± 0.015	-	0.725 ± 0.015	-
DR9 (It. Opt.)	0.278 ± 0.013	13%	0.725 ± 0.015	0%

that it is possible to gain in accuracy in two independent ways and the combination of both methods is expected to provide even better constraints on cosmological parameters.

4.3. DR9

Following the same procedure as for the DR7 LRG sample, we computed the spherically averaged 2-point correlation function using the Landy-Szalay estimator and the iterative optimal estimator. The results as shown in the right panel of Fig. 11. The corresponding values of α are given in Tab. 2.

We see a clear improvement on the precision of the α measurement with respect to the Landy-Szalay one. The values are in agreement, but the iterative estimator gives us 28% more accurate result.

As discussed for the DR7 data, $\alpha_{\rm Measured}$ and its error for DR9 CMASS data are consistent with the measurements with PTHalos mocks, as can be seen in Fig. 12 (right) and 13 (right).

The BOSS DR9 CMASS result [Anderson et al., (2012)] using the correlation function only is $\alpha = 1.016 \pm 0.017$ and $\alpha = 1.024 \pm 0.016$ before and after reconstruction respectively. However in the case of DR9, the improvement of 6% due to reconstruction is much lower that the one expected with the new iterative estimator. Meanwhile, this result is consistent with our values with both estimators (Tab. 2) well within 1-σ.

5. Cosmological constraints

The improvement on cosmological parameter constraints using the iterative optimal estimator is illustrated in Fig. 14. These constraints are obtained using a Monte Carlo Markov Chain within an open ΛCDM cosmology using CMB data only. The chain[3] was re-sampled with our BAO α constraints. The marginalized constraints on Ω_m and Ω_Λ are given in Tab. 3.

The overall gain on the cosmological parameters is between 13 and 22% (except for Ω_Λ for DR9). With the iterative optimal estimator applied on DR7 data, the accuracy on Ω_m and Ω_Λ is comparable to that measured with the Landy-Szalay estimator applied to the DR9 sample, even though the DR9 has a density that is three times larger and twice the volume of DR7.

6. Conclusions

We have designed a new two-point correlation function estimator, which is a linear combination of all possible ratios (up to second order) of pairs counts between data and random samples. The linear combination can be optimized to minimize the variance of the correlation function for a given geometry. We developed an iterative procedure to make this new estimator independent of the cosmology of the simulated data used in its optimization. We have shown on lognormal, second-order perturbation theory and N-body simulations that the decrease in size of the correlation function error bars is around 25%, relative to the well known Landy-Szalay estimator. The improvement is not mitigated by extra correlations in the covariance matrix of the two-point correlation function.

This result is not contradictory with the fact that the Landy-Szalay estimator was shown to be of minimal variance, since this is true only for a vanishing correlation function and a simple geometry. Current galaxy surveys do measure a nonzero correlation function even on large scales and they have quite complex geometry.

Finally, we have applied our method to SDSS DR7 and DR9 data, achieving an improvement of 10-15% on the value of the cosmological parameters Ω_m and Ω_Λ. We achieve a similar accuracy with our estimator on the DR7 sample as with the Landy-Szalay estimator on the DR9 sample.

Our method can be easily applied to any dataset and requires modest extra CPU time, as any analysis anyway requires a large number of simulated catalogues to test for systematic effects and to estimate the covariance matrix.

For future developments, we would use Principal Component analysis to identify the combination of ratios that contributes most to minimize the correlation function variance. The optimization could then be limited to the most relevant combinations.

This method can be easily extended to the study of the anisotropic correlation function. The coefficients would be optimized to simultaneously minimize the variance of the monopole and quadrupole. That approach would produce better constraints on redshift space distortions [Kaiser (1987)] physical parameters and on the Alcock-Paczynski test [Alcock and Paczynski (1979)].

The optimized iterative estimator could be easily applied to mark correlation functions as well [e.g., Skibba et al. (2006); Martinez et al. (2010)].

Acknowledgements. We would like to thank the SDSS-III collaboration for such wonderful data. We thank N. Padmanabhan and J.K. Parejko for making their kd-tree code available.

We used the "gamma" release LRG galaxy mock catalogs produced by the LasDamas project; we thank the LasDamas collaboration for providing us with this data.

We would like to thank R. Skibba, Chia-Hsun Chuang, Lado Samushia and Graziano Rossi for helpful suggestions and comments.

This project was supported by the Agence Nationale de la Recherche under contract ANR-08-BLAN-0222.

[3] The MCMC from WMAP7 is available at `http://lambda.gsfc.nasa.gov/`.

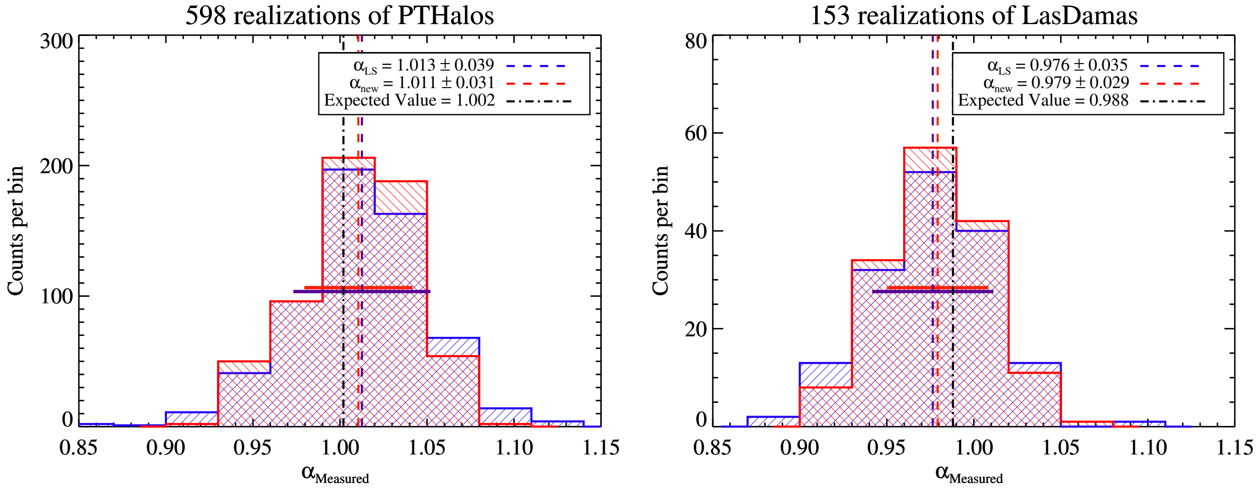

Fig. 10. Histogram of α_{Measured} for the PTHalos (left) and LasDamas (right) realizations using the Landy-Szalay (blue) and the iterative optimal estimators (red). The average values over the realizations, shown in the legend, are represented as vertical dashed lines of the same color, and their error as horizontal bars. The expected value is shown as a black dot-dashed vertical line. (Coloured version of the figure is available online)

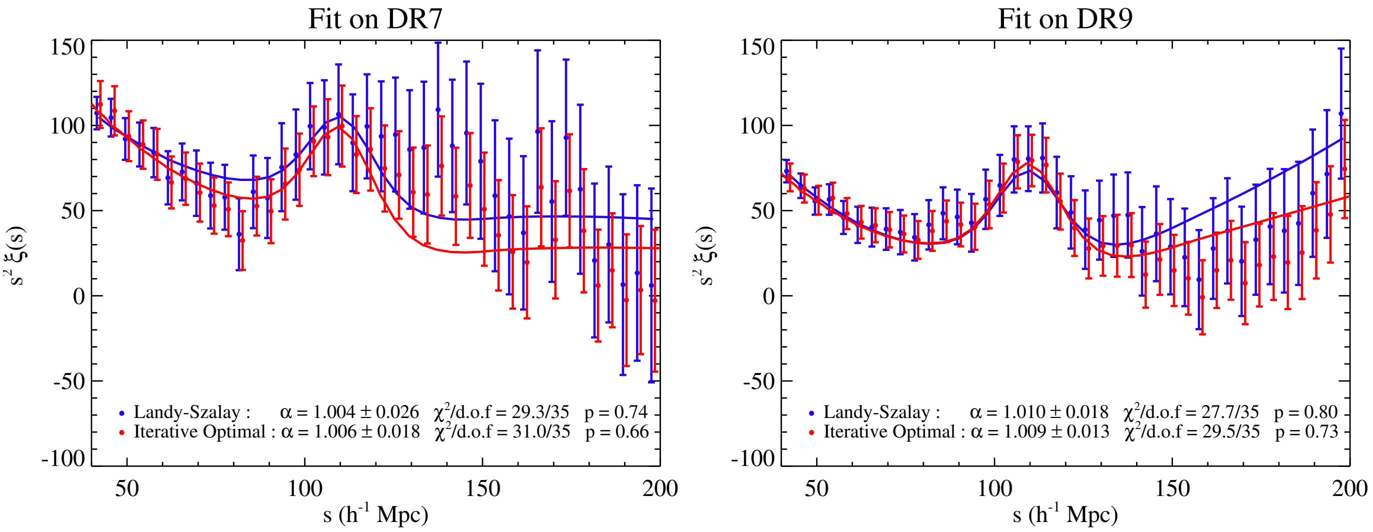

Fig. 11. Correlation functions obtained for DR7 LRG (left) and DR9 CMASS (right) data samples using the Landy-Szalay (blue points) and the iterative optimal estimator (red points). Their best fit is shown by solid lines and α_{Measured} is given in the legend together with the χ^2/d.o.f. and its probability. The covariance matrices used for these fits for each iteration of the fit are based upon the LasDamas and PTHalos mocks covariance matrices, respectively, also obtained using both estimators. The error bars are the square root of the diagonal elements of these matrices. (Coloured version of the figure is available online)

Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III web site is http://www.sdss3.org/.

SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, University of Cambridge, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofisica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University.

References

Abazajian, K. N. et al., ApJS, 182, 543, 2009
Ahn, C. et al., ApJs, submitted, arXiv:1207.7137S
Alcock, C., & Paczynski, B., 1979, Nature, 281, 358
Aihara, H., et al. 2011, ApJS, 193, 29
Anderson, L. et al., submitted, `arXiv:1203.6594`
Blake, C. at al. MNRAS, 415, 1707, 2011
Bolton, A., et al, AJ, 144, 144, 2012.
Cabré, A & Gaztañaga, E., arXiv:0807:2460v3
Coles, P. and Jones, B. MNRAS, 248, 1-13, 1991
Davis, M. and Peebles P.J.E., ApJ, 267, p465, 1983
Dawson, K. et al., AJ, submitted.

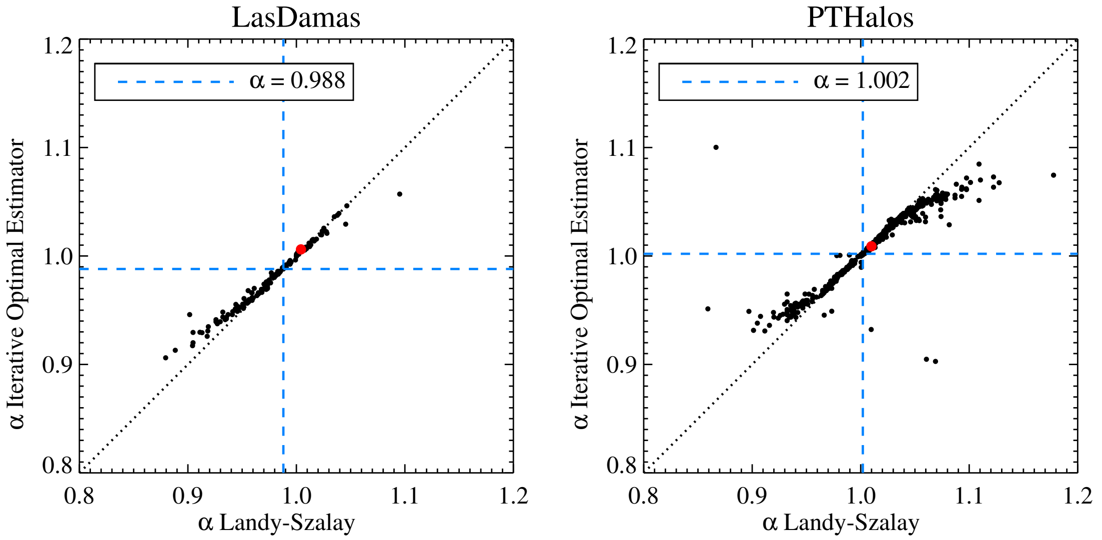

Fig. 12. Comparison of $\alpha_{\rm Measured}$ using the Landy-Szalay and the iterative optimal estimators for the mocks (small black points) and the real data (red point) for LasDamas and DR7 (left), and PTHalos and DR9 (right). The expected values of α are shown by dashed blue lines. (Coloured version of the figure is available online)

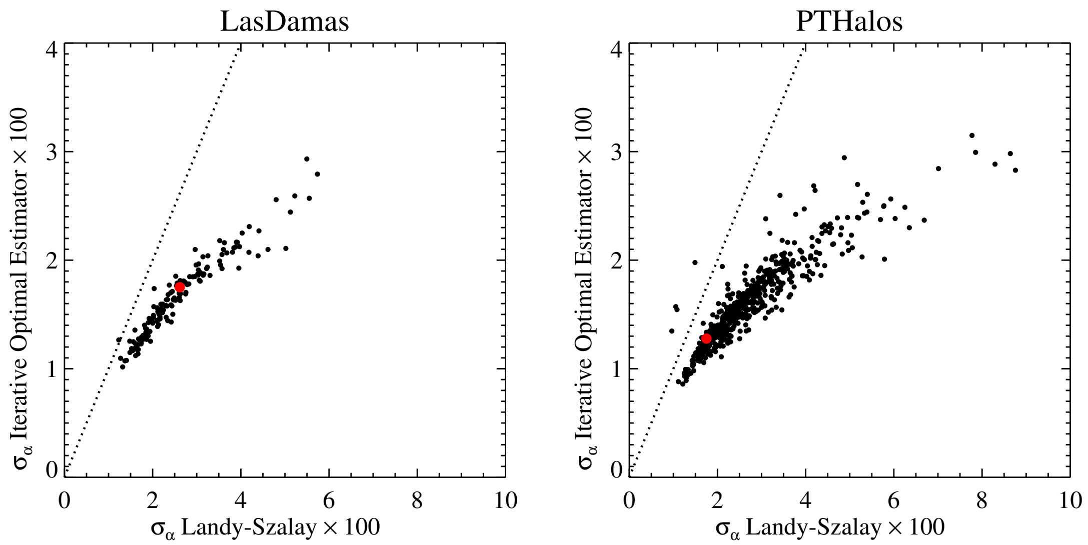

Fig. 13. Comparison of the error on $\alpha_{\rm Measured}$ using Landy-Szalay and the iterative optimal estimators for the mocks (small black points) and the real data sample (red points). As in Fig. 12, the left plot is for LasDamas and DR7 LRG data and the right one for PTHalos and DR9 data. The black dotted line corresponds to the same error for the two estimators. (Coloured version of the figure is available online).

Eisenstein, D. J., Hu, W., 1998, ApJ, 496, 605
Eisenstein, D. et al., Astron. J., 122, 2267, 2001
Eisenstein, D. J., et al. 2005, ApJ, 633, 560
Eisenstein, D. J., Seo, H.-J., Sirko, E., & Spergel, D. N., 2007, ApJ, 664, 675
Eisenstein, D.J. et al., Astron.J.142:72,2011, `arXiv:1101.1529`
Efstathiou, G., Sutherland, W. J., & Maddox, S. J. 1990, Nature, 348, 705
Fukugita et al., Astronomical Journal v.111, p.1748 (1996)
Granett B.R. et al., Astrophys.J. 701 (2009), `arXiv:0812.1025`
Gunn, J.E., et al. , AJ, 116, 3040 , 1998.
Gunn, J.E., et al., AJ, 131, 2332 , 2006.
Hamilton, A.J.S., Astron. J. v417, p19, 1993
Hewett, P.C., MNRAS, 201, p867, 1982
Kaiser, N., 1987, MNRAS, 227, 1
Kazin E., et al., 2010, ApJ, 710, 1444
Kerscher (1999), A&A, 343, 333.
Kerscher, Szapudi and Szalay, 2000, ApJ, 535, 13.
Komatsu, E., et.al., 2011, ApJS, 192, 18, `arXiv:1001.4538`
Landy, S. D. & Szalay, A. S. 1993, ApJ, 412, 64
Lewis, A., & Bridle, S., 2002, Phys. Rev. D, 66, 103511
Liddle, A. and Lyth, D., Cosmological Inflation and Large Scale Structure, Cambridge University Press (2000)
Manera M., et al., 2012, MNRAS submitted, `arXiv:1203.6609`
Martinez et al. 2010, A&A, 513, A22.
McBride C. et al., in preparation
Mehta K.T. et al., MNRS submitted, `arXiv:1202.0092v2`
Padmanabhan, N. et al., in preparation, 2012b
Padmanabhan et al. 2007, MNRAS, 376, 1702
Peebles, P.J.E. and Hauser, M.G., ApJS, 28, p19 (1974)
Percival, W.J., et al., 2007, MNRAS, 381, 1053
Perlmutter, S., et al. 1999, ApJ, 517, 565
Reid et al, MNRAS, Volume 426, Issue 4, pp. 2719-2737.
Pons-Bordería et al. 1999, ApJ, 523, 480.
Riess, A. G., et al. 1998, AJ, 116, 1009
Sanchez, A. G., et al., 2012, MNRAS submitted

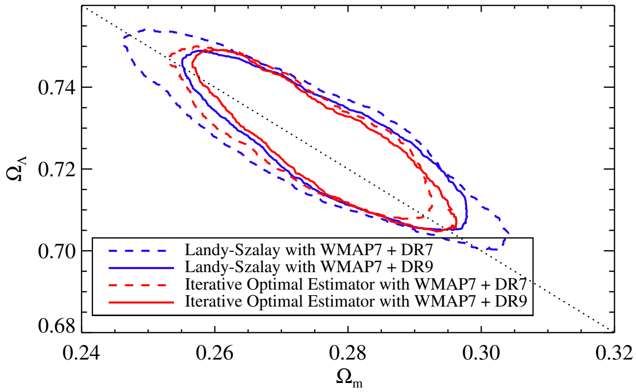

Fig. 14. The 68% joint constraints in the $(\Omega_m, \Omega_\Lambda)$ plane for an open ΛCDM cosmology combining CMB (WMAP 7 years [Komatsu et al., (2011)]) and either DR7 (dashed lines) or DR9 (solid lines) SDSS BAO data, with either the Landy-Szalay estimator (blue) or the iterative optimal estimator (red). (Coloured version of the figure is available online).

Scoccimaro, R. and Sheth, R.K., MNRAS, 329, 629, 2002
Schlegel, D. et al. 2012, in preparation
Slosar, A. et al., JCAP, Issue 09, pp. 001 (2011)
Smee, S.A., et al. , AJ, submitted.
Sherwin, B.D. et al., PRL 107, 021302 (2011), arXiv:1105.0419
Skibba et al.,2006, MNRAS, 369, 68.
Weinberg, D.H. et al., Physics Report, 2012 arXiv:1201.2434
Xu, X., Cuesta, A. J., Padmanabhan, N., Eisenstein, D., McBride, C. K., 2012, arXiv:1206.6732
York, D.G., et al, AJ, 120, 1579, 2000.